A feature-integration theory of attention

Experimental report on ″the pop-out effect in visual attention experiments″

Frederic Becker and Pia Scholz

April 3, 2022

A project paper presented for the course
″Introduction to reproducible research practices via browser-based replication″
WS 2021/2022

Department of Linguistics
University of Tuebingen
Prof. Dr. Michael Franke

Abstract

The feature-integration theory of attention has over 15000 citations and is one of the most influential theories in visual information processing (Quinlan, 2003; Wolfe, 2020). The postulated idea is that focal attention is crucial for the correct perception of objects characterised by multiple features, such as color and shape. Visual search experiments conducted by A. M. Treisman and Gelade (1980) form the empirical foundation of this theory. In this reaction time experiment participants are asked to respond whether or not a certain letter is within a set of presented letters. It was demonstrated that search times for letters that were characterized by a combination of features increased linear with the number of letters presented, whereas this was not the case if the letter was defined by a single unique feature such that the letter of interest popped out. We present a replication of this study. Indeed, we found that the search for letters with conjuncted features strongly depended on the number of presented letters. In contrary to the original study, we found a small but significant linear relation between the number of presented letters and response times if the feature popped out.

The material for this study can be found at our Github Repository.
Our experiment can be accessed here.

1 Introduction

Every time we take a look at our surroundings, we instantly grasp them and form an impression of the objects and elements that surround us. This cognitive process integrates basic visual perceptions to a rich semantic meaningful representation. The hereby identified objects are usually complex in the sense that they are defined by various properties (e.g. color and shape). The debate on how we acquire such impressions was historically dominated by two theories. On the one hand, the Gestalt streaming assumed that properties are recognised only after an initial recognition of the object itself (Monahan & Lockhead, 1977). Such a process first recognises an entity which is then decomposed into properties. On the other hand, associationists suggested that a complex impression is inherently built upon a combination of elementary sensations. In this, properties are first recongized which are then combined to entities. The debate was strongly influenced towards the associationistic view by the work of Anne Treisman and Garry Gelade on the feature-integration theory of attention (A. Treisman, 1977).

The feature-integration theory of attention postulates that features such as orientation, brightness and colour are processed automatically, fast and in parallel to be arranged in separate spatial representations. Each of these representations depict how a certain feature is distributed over our visual field. In a subsequent processing stage focal attention acts as glue to bind features of different type to a unitary object based on their spatial overlap. Focal attention thus plays a central role with its exclusive ability to map multiple features into an object. Focal attention is considered a very limited cognitive resource that can only be focused on one point at a time. As a consequence the process of feature-integration can only occur in a serial fashion, however once an entity is formed it persists in perception. The central claim of the feature-integration theory of attention is thus that attention is crucial for the correct perception of objects characterised by multiple features (**?**).

Treisman and Gelade examined that assertion with a visual search experiment. In this task, different letters are presented on a screen. The participant has to report as fast and as accurate as possible if a specific letter, called target, is within the presented letters or not. Treisman and Gelade found that in case the target had to be identified by multiple features search times increased linearly with the number of letters presented. In this setting, the features of the target (blue and X) partially overlapped in both domains (shape and color) with the features of the distractors (blue and T or read and X). To correctly identify the target it was thus necessary to search for the correct combination of features, and to make use of a feature integration process. The authors interpreted the linear dependence between the number of letters and the search time as evidence that the feature binding process is indeed serial. However, search times did not increase linearly when the target could be identified by a unique feature that made the target letter "pop out" from the rest of the letters. In this setting, one feature of the target (green or S) did not overlap with the features of the distractors (blue and T or red and X). To correctly identify the target it was thus only necessary to search for a single feature. The feature-integration theory of attention suggests that this process should not be limited by focal attention and can thus occur parallel, which was indeed confirmed by the the static result pattern.

Treisman and Gelade's work is cited over 15000 times and marked the starting point of the rise of the feature-integration theory of attention to one of the most influential theories in visual information processing (Quinlan, 2003; Wolfe, 2020). Due to the major influence of the feature-integration theory on our current scientific knowledge it is essential that it is founded on solid empirical evidence. Empirical evidence gains its credibility from the fact that the core observation can be replicated by following the methods and procedure that provoked it in the first place (Zwaan, Etz, Lucas, & Donnellan, 2018). Without the property of reproducibility we would not be able to distinguish random from systematic observations and importantly only the latter allows for scientific inference. A series of attempts to replicate well-known psychological studies have partially failed, and led to a severe methodological debate, known as the 'replication crisis' (Ioannidis, 2005). The causes of replication failure are diverse and include under-powered original studies, publication bias, excessive researcher degrees of freedoms and p-hacking (Ioannidis, 2005; Munafò et al., 2017; Zwaan et al., 2018). To restore reproducibility, a number of methodological corrections have been proposed, the concept of preregistration being the best known example. Although improved research standards provide the basis for adequate replication rates, there is no alternative to conducting replication studies on a large scale to verify scientific findings (Zwaan et al., 2018).

This seminar paper reports a replication study of 'Experiment 1' also known as the 'pop-out effect' as described in A. M. Treisman and Gelade 1980. We found a similar result pattern to the one presented in the original study, even though the core claim that search times in the 'pop-out' condition are independent of the number of presented letters did not hold our statistical analysis. In the next section we will report the methods and procedure which is followed by the results. In section 4 we will summarize and conclude our findings.

2 Methods

Participants

The 15 subjects ($range_{age}$ = 20-56 years; M_{age} = 27,133) were to equal parts male (N_{male} = 7) and female (N_{female} = 7). Five subjects stated to have graduated high school, another five have graduated college and four have earned a higher degree. One subject has neither stated age, nor gender or educational level. At least two participants had previously taken part in the pilot study described in the preregistration report (to be found in our Github Repository).

All participants were recruited personally or text-based via a messenger client (e.g. WhatsApp). Each participant received a link to the experiment and a short (1-2 sentences) explanation on the procedure. They were told to only perform the experiment on a laptop/desktop computer and to use corrective eye wear if needed.

The participants did not receive any kind of compensation.

There were four additional subjects recruited in the same way as described above. They provided pilot rating values: two male and one female subject with $range_{age}$ = 22-56 and M_{age} = 25,5. One subject chose not to share age, gender or education.

Materials

This visual search task was adapted from A. M. Treisman and Gelade 1980.
In the experiment we made use of seven distinct items. Each item was a letter (T, X or S) combined with a colour (green, brown or blue). Out of these, T_{brown} and X_{green} were used as distractors that appeared regardless of trial condition. The remaining five letters T_{green}, T_{blue}, X_{blue}, S_{brown} and S_{green} were the targets.
Just like A. M. Treisman and Gelade 1980 anticipated we made use of three factors that were manipulated within-subjects. The experiment was divided into two **conditions**. The conjunction condition showed a target that combines features (T_{green}) whereas the disjunction condition includes only targets defined by a single distinct feature (T_{blue}, X_{blue}, S_{brown} and S_{green}). Whether or not the target letter was present was determined by two **trial types**: positive and negative. In positive trials a single target letter and in negative trials no target letter was shown. The third factor we manipulated was **display size** which refers to the number of items shown (1, 5, 15 or 30). In positive trials one target letter and an equal number of both distractors could be seen. For trials with an odd number of distractors two variants were created. In each one of the two distractors dominated the other by the count of one. All together, we ended up with 16 blocks á 8 stimuli cards. A more detailed overview can be found in our repository.
The 128 stimuli cards were created semi-automatically in R and automatically saved as .png files. Distractors and targets were scattered randomly across the card. We chose a grid size of 7x7 items which we used for all display sizes. By doing this we could keep the distance between neighbouring letters constant regardless of how densely the cards were packed. Also, fovea distance could be kept at a constant by doing so.

Procedure

The experiment itself started in a new tab with a brief informational texts to give participants an idea on how the experiment proceeds and to inform about the kind of data we were planning on collecting. Participants were then asked to adjust their screen size to a grey rectangle by making use of a standard credit card and the zoom in/out function. We asked for this adjustment because we wanted to keep fovea distance at a constant for as good as possible. This first part was followed by 4 blocks of the experimental task and finally a short demographic survey. Each block was exclusively part of the conjunction (C) or the disjunction (D) condition and consisted of 128 trials (16 positive and negative per display size). The block sequence was randomised (CDCD and DCDC) as well as the sequence of trials within each block. Half of the participants were shown the block sequence CDCD, the other half the DCDC sequence. Before each block a short instruction on the target letter was given. For the disjunction condition the instruction was: "S or any blue letter" and for the conjunction: "green T".

Every trial started with a central fixation cross displayed for one second, followed by the search configuration. After response, the participant is shown the measured reaction time in milliseconds and the correctness of response. Participants were instructed before to response as fast and accurate as possible. For reporting a positive trial the button 'J' and for reporting a negative trial the button 'F' had to be pushed. In case of erroneous trials, the consecutive trial was defined as a "dummy" trial, i.e. no data was recorded for this trial. In addition, the erroneous trial was repeated later in the same block. After completing the four blocks we conducted a short demographic survey asking for age, gender and educational level of the participant. As soon as they received the message that all data has been successfully uploaded participants were free to close the tab.

Data analysis

We used the statistical analysis program R. For Bayesian regression we used the model brms (Bürkner, 2021) and the default (flat) priors of the 'brms' package for the effect coefficients. The alpha level was 0.05. Hypothesis were tested by evaluating if the 95% credible intervals (CI) contained the value suggested by the null hypothesis. We excluded all trials in which the response time was faster than 200 ms or slower than 3000 ms. For the analysis of response times we additionally excluded trials with incorrect responses. The analysis was preregistered. We will report estimates for the slopes (β) and the corresponding Credible Intervals (CI).

3 Results

Visual search in the conjunction condition

The mean response in the conjunction condition was 1134 ms (SD = 407). The response times increased linearly with display size, as illustrated in Figure 1. This was the case for negative (slope: $\beta = 27.8,\ CI = [25.4, 30.1]$) and positive trials (slope: $\beta = 16.9,\ CI = [14.5, 19.3]$). Further, the slope of search time was significantly steeper in negative trials than in positive trials of the conjunction condition (negative minus positive slope: $\beta = 10.8,\ CI = [7.9, 13.7]$). The linear influence of display size almost halved in positive trials (the ratio of the slopes was 0.61). As hypothesized and proposed by A. M. Treisman and Gelade (1980), the results suggest that visual search in the conjunction condition is serial and self-terminating.

Visual search in the feature condition

The mean response in the feature condition and 982 ms (SD = 334). According to our hypothesis, we found that the response times in the feature condition increased linearly with display size in case the target was absent (slope: $\beta = 14.6,\ CI = [12.3, 16.9]$). In contrary

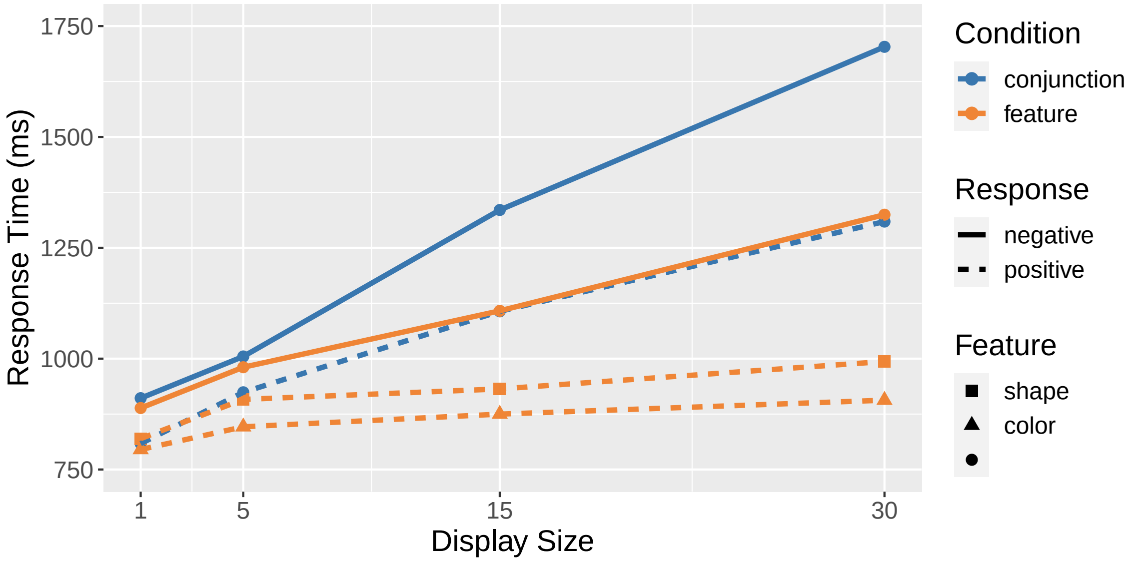

Figure 1: Mean search times per condition, display size and response. In the feature condition (orange), the target was defined by either a unique color or unique shape, whereas in the conjunction condition (blue) the target was defined by a unique combination of these. Both conditions included trials in which the target was present (dotted line) and trials in which the target was absent (solid line).

to our hypothesis, we found that this was also the case when the target was present (slope: $\beta = 4.3,\ CI = [1.8, 6.6]$). However, the slope was very small and it is in the range of the original estimate obtained by A. M. Treisman and Gelade (1980) (see Table 1). These results indicate that search times in the feature condition depend on display size, with only small dependence in positive trials. A preregistered exploratory analysis estimated the slopes for response times in which the target was present and described by a unique color to be independent of display size (color slope: $\beta = 2.4,\ CI = [-1.59, 6.52]$). This was also the case for trials those targets were described by shape (shape slope: $\beta = 3.4,\ CI = [-0.69, 7.60]$). Although a tendency towards dependency can be recognized.

Visual search with only one item

The response times in trials with only one letter did not significantly differ between conditions($t(480) = 0.46, p = .64$). The mean response time was 858 ms (SD = 201) in the conjunction condition and 849 ms (SD = 190) in the feature condition. This is consistent with our hypothesis and the implications of the feature-integration theory of attention, since serial and parallel search should proceed in the same way in this setting.

Error rates

The error rate was 5.7% in the conjunction condition and 5.3% in the feature condition. Against our hypothesis, that there should be no effects of display size on error rates, we found that the error rate for positive trials in the conjunction condition increased with display size (slope $\beta = 0.08$, $CI = [0.05, 0.11]$). This may be related to the increasing degree of complexity introduced by display size.

Table 1: Estimates of the linear regression coefficients. The asterisk marks significant deviation from linearity.

		This work		Treisman et al. (1977)	
Condition	Response	Intercept	Slope	Intercept	Slope
conjunction	positive	821	16.9*	398	28.7*
	negative	884	27.8*	397	67.1*
feature mean	positive	829	4.3*	448	3.1
	negative	889	14.6*	514	25.1*
feature color	positive	811	2.4	455	3.8
feature shape	positive	849	3.4	441	2.5

4 Discussion and Conclusions

This replication study of the 'pop-out effect' (A. M. Treisman and Gelade 1980) found a similar result pattern as the original study. We could replicate the finding that the visual search in the conjunction condition is serial and self-terminating. However, we could not replicate the finding that search times in the feature condition do not increase linearly with display size. Still, this dependency was very small suggesting that the visual search in the feature condition was not strongly affected by display size, allowing the conclusion that the search mechanism is potentially nearly parallel. The significant differences between the search for objects defined by a conjunction of features and objects defined by a unique feature, support the central claim of the feature-integration theory of attention that focal attention is a central processing component.

The divergence of the result pattern in the feature condition may be related to the fact that the number of trials in the original study was higher, which could have led to training effects encouraging faster response times especially in trials with a larger display size. Another, major difference is that the original study was conducted under controlled conditions in a laboratory, whereas this study was browser-based and thus less controlled. This could also be the cause for the about twice as large Intercepts of the response time models in our study compared to the original study (see Table 1). Overall the replication was designed to match the original study in all central aspects; we used the same letters, feature configurations and display sizes, and employed a similar procedure and analysis. A potential improvement is to test a wider range of display sizes to obtain a more robust estimate of the relationship between response times and display sizes.

Conclusively, the result pattern presented in A. M. Treisman and Gelade (1980) could be replicated, with the exception of steady response times for the pop-out search.

References

Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software*, *100*(5), 1–54. doi: 10.18637/jss.v100.i05

Ioannidis, J. P. A. (2005, 08). Why most published research findings are false. *PLOS Medicine*, *2*(8), null. Retrieved from `https://doi.org/10.1371/journal.pmed.0020124` doi: 10.1371/journal.pmed.0020124

Monahan, J. S., & Lockhead, G. R. (1977). Identification of integral stimuli. *Journal of Experimental Psychology: General*, *106*(1), 94-110. Retrieved from `https://doi.org/10.1037/0096-3445.106.1.94` doi: 10.1037/0096-3445.106.1.94

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., ... Ioannidis, J. P. A. (2017, Jan 10). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 0021. Retrieved from `https://doi.org/10.1038/s41562-016-0021` doi: 10.1038/s41562-016-0021

Quinlan, P. T. (2003). Visual feature integration theory: past, present, and future. *Psychological bulletin*, *129*(5), 643.

Treisman, A. (1977, Jan 01). Focused attention in the perception and retrieval of multidimensional stimuli. *Perception & Psychophysics*, *22*(1), 1-11. Retrieved from `https://doi.org/10.3758/BF03206074` doi: 10.3758/BF03206074

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97-136. Retrieved from `https://www.sciencedirect.com/science/article/pii/0010028580900055` doi: https://doi.org/10.1016/0010-0285(80)90005-5

Wolfe, J. M. (2020, Jan 01). Forty years after feature integration theory: An introduction to the special issue in honor of the contributions of anne treisman. *Attention, Perception, & Psychophysics*, *82*(1), 1-6. Retrieved from `https://doi.org/10.3758/s13414-019-01966-3` doi: 10.3758/s13414-019-01966-3

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*, e120. doi: 10.1017/S0140525X17001972